INFO6250 Ranking System Project Report

By

Ravi Kiran (001491808)
Veena Iyer (001447061)

1. Introduction:

A sports ranking system is a system that analyzes the results of sports competitions to provide ratings or each team. Ratings are numerical representations of competitive strength, often directly comparable so that the game outcome between any two teams can be predicted.

Rankings can be directly provided sorting each team's ratings and assigning an ordinal rank to each team, so that the highest rated team earns the #1 rank. Rating systems provide an alternative to traditional sports standings which are based on win-loss-tie ratios.

2.Problem Statement:

We implemented Ranking Algorithm for teams in English Premier League. It is a ranking problem which will create ranks for the teams of English Premier League by

taking input as team names and the results of each team with every other team and also taking present environment of the COVID19. We have considered two seasons where in the current season has matches which could not be conducted due to COVID19.

Thus we have used Colley Rankings- Mr. Wesley N. Colley suggested that instead of calculating rankings based on win-loss ratio like

ratings = [total wins] / [total games], ratings should be calculated based on following formula.

3. Detail Description:

Why do we need Ranking System?

Will a simple win-loss ratio not be enough? Consider there is a tournament with 4 teams playing in it and following are the results, and final table of the tournament standings so far.

Rank	Team	Win	Loss
1	Manchester United	6	0
2	Chelsea	3	3
3	Arsenal	3	3
4	Liverpool	0	6

As you can see, Manchester United has won all of their 6 games and sit comfortably on top, Liverpool have lost all their games and lie on bottom of table. Arsenal and Chelsea share the spoil having been defeated by Man United in both rounds, defeating Liverpool in both rounds and then splitting the tie between themselves with Arsenal winning in 1st round and Chelsea returning the favor in reverse fixture.

Now imagine we begin 3rd round of fixtures and first couple of game's results were as below.
Game 1: Arsenal defeat Manchester United
Game 2: Chelsea defeat Liverpool

With this the table will look like below-

Rank	Team	Win	Loss
1	Manchester United	6	1

2	Chelsea	4	3
3	Arsenal	4	3
4	Liverpool	0	7

Now wait a minute, does Chelsea and Arsenal really deserve to get same rankings? Isn't it pretty obvious that Arsenal has defeated Manchester United who were dominant team in the tournament so far, whereas Chelsea defeated Liverpool who had anyways lost all their games. Definitely Arsenal should get more credit and should be ranked higher than Chelsea. But how do we do that? If we are using just a simple win-loss ratio, there is no way to differentiate them.

Thus instead of calculating rankings based on win-loss ratio like
ratings = [total wins]/[total games], we use
ratings = [1+total wins]/[2+total games]
If you look at this formula, before start of any tournament, all teams will begin with a rating of 0.5, as both 'total wins' and 'total games' will be zero
We can thus write total-wins as

$$
\begin{aligned}
total-wins &= \frac{total-wins}{2} + \frac{total-wins}{2} \\
&= \left[\frac{total-wins}{2} + \frac{total-wins}{2}\right] + \left[\frac{total-loss}{2} - \frac{total-loss}{2}\right] \\
&= \left[\frac{total-wins}{2} - \frac{total-loss}{2}\right] + \left[\frac{total-wins}{2} + \frac{total-loss}{2}\right] \\
&= \left[\frac{total-wins}{2} - \frac{total-loss}{2}\right] + \frac{1}{2}[total-wins+total-loss] \\
&= \left[\frac{total-wins}{2} - \frac{total-loss}{2}\right] + \frac{1}{2}[total-games]
\end{aligned}
$$

Looking closely into ½ * [total-games] term for a single team, we can say that ½ * [total-games] = ½ * [1+1+1…1] where '1' stands for a game played against each opponent. This in turn can be written as ½ * [total-games] = [1/2 + ½ + ½+..]

Thus rewriting the ratings as

$$ratings \approx \left[\frac{1 + (total - wins)}{2 + (total - games)} \right]$$

Expanding total-wins and writing down rating for individual team, we can write as.

$$r_i = \frac{1 + [w_i - l_i]/2 + \sum r_j}{2 + t_i}$$

Where r_i, w_i, l_i, and t_i represent rating, wins, loss and total game for the team of interest and $\sum r_j$ is sum of opponent's ratings.

Now lets consider a case of two teams. As stated above, before beginning, they both will have a rating of 0.5 each. Now let's suppose they played one game against each other. Let rW and rL represent new ratings for winning and loosing team respectively for both teams, and consider their w_i and l_i values, we will get following results from equation (A)

$$3r_W - r_L = 3/2$$
$$-r_W + 3r_L = 1/2,$$

This represents a simple two variable linear system, which when solved gives us our new ratings as rw = 5/8 and rL = 3/8. For 2 teams, equation (A) resulted in two variable linear system, for N teams, it will result in N variable linear system.
In data structure world when we want to solve N variable linear equations we use matrix. So rewriting matrix as

$$C\vec{r} = \vec{b},$$

4. Implementation Details

We have considered data set of two season, current season affected by COVID19 and previous season. Thus we check for each team against every other team. We will have in the matrix form one team coming across itself and this scenario needs to be handled. This element in the matrix is filled as 2 + total-games for corresponding team.

Then each team playing another team should be replaced by negative value of the number of games played between team represented by row and column for that cell. So for example cell intersecting Man United row and Chelsea column should have a number equal to negative value of number of Man United vs Chelsea games. And like that we will fill rest of the cells in matrix. Once we solve this N variable linear equation $Cr = b$, we will get vector value for 'r', which represents ratings for each individual team.

We thus use one *two dimensional matrix* to read the excel file which has details of all teams and their wins, loss or draws. We use an *array* to store the wins and loss as a result for each team. In cases of draws we divide the draws into equal halves and add them to win and loss of each team to balance the probability. We also use *list* to store the names of teams. As is the rule of EPL the bottom three from a particular season are removed from preforming in the next season and three new teams are added to the new season.

Based on season selected we have a *HashMap* which validates user input and calculates the probability of win for each of them in a match. Now taking current scenario of COVID19 we have many matches that were canceled or pushed for later and these scores in matrix are set to n/a which is then set to 0 in our calculation as there has been no match and thus should not affect the win or loss outcome. The probability for these teams are calculated on the basis of performance previous season and a probability is suggested. These ratings are then calculated between the two dimensional array which has the points between one team and corresponding team and also the individual team results in one dimensional array. Using jama matrix to solve these arrays we get ratings for each team in the form of a matrix. This is then used to populate *HashMap* with the team name from the list and the result from the matrix. We use LinkedHashMap to sort map based on values and display the rankings.

```
        Matrix ans = lhs.solve(rhs);

        Map<String, Double> teamRanks = new HashMap<>();
        for(int i=0;i<teamNames.length;i++) {
            if (teamNames[i] != null)
                teamRanks.put(teamNames[i], ans.get( i: i - 1, i1: 0));
        }

        LinkedHashMap<String,Double> sortedByValue = new LinkedHashMap<>();
        teamRanks.entrySet().stream().sorted(Map.Entry.comparingByValue(Comparator.reverseOrder()))
                .forEachOrdered(x -> sortedByValue.put(x.getKey(), x.getValue()));

        System.out.println("");
        System.out.println("################################");
        System.out.println("EPL Team Rankings are as below");
        System.out.println("################################");


        for(Map.Entry<String,Double> sorted : sortedByValue.entrySet()){
            System.out.println(sorted.getKey() + " = " + sorted.getValue());
        }
    }
}
```

Code for sorting the rankings based on ratings

5. Program Output

The below are console output as the program executes.

Select which season you need to view the rankings for – dataset selection

```
We have two datasets - the 2018-2019 dataID-19
 Which seasons ranking would you like to view?
 Enter 1 for 2018-2019 and 2 for COVID-19 dataset
|
```

```
We have two datasets - the 2018-2019 dataID-19
 Which seasons ranking would you like to view?
 Enter 1 for 2018-2019 and 2 for COVID-19 dataset
2

EPL fan? Check which team wins and their Rankings
Enter --1 for -- Arsenal
Enter --2 for -- Aston Villa
Enter --3 for -- Bournemouth
Enter --4 for -- Brighton & Hove Albion
Enter --5 for -- Burnley
Enter --6 for -- Chelsea
Enter --7 for -- Crystal Palace
Enter --8 for -- Everton
Enter --9 for -- Leicester City
Enter --10 for -- Liverpool
Enter --11 for -- Manchester City
Enter --12 for -- Manchester United
Enter --13 for -- Newcastle United
Enter --14 for -- Norwich City
Enter --15 for -- Sheffield United
Enter --16 for -- Southampton
Enter --17 for -- Tottenham Hotspur
Enter --18 for -- Watford
Enter --19 for -- West Ham United
Enter --20 for -- Wolverhampton Wanderers
```

On selecting 2 as the dataset to view we are shown teams. To avoid human error provided functionality to enter numbers.

```
16 1
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.dom4j.io.SAXContentHandler (file:/home/batman/.m2/repository/dom4j/dom4j/1.6.1/dom4j-1.6.1.jar) to method com.sun.org.apache.xerces.internal.parsers.AbstractSAXParser$LocatorProxy.getEncoding
WARNING: Please consider reporting this to the maintainers of org.dom4j.io.SAXContentHandler
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
#####################################
Home Team, Away Team : Result(home team)
#####################################
Manchester United,Crystal Palace : L
Southampton,Manchester City : L
Watford,Bournemouth : D
Chelsea,Manchester City : W
Wolverhampton Wanderers,Brighton & Hove Albion : D
Newcastle United,Crystal Palace : W
Bournemouth,Manchester United : W
Tottenham Hotspur,Newcastle United : L
West Ham United,Manchester City : L
Liverpool,Sheffield United : W
Aston Villa,Norwich City : W
Liverpool,Leicester City : W
Watford,Everton : L
Chelsea,Aston Villa : W
Leicester City,Southampton : L
West Ham United,Crystal Palace : L
Norwich City,Aston Villa : L
Watford,West Ham United : L
Crystal Palace,West Ham United : W
Bournemouth,Sheffield United : D
Leicester City,Sheffield United : na
Burnley,Manchester United : L
Manchester United,Newcastle United : W
Crystal Palace,Chelsea : L
Watford,Arsenal : D
Watford,Burnley : D
Everton,Crystal Palace : W
Manchester United,Manchester City : W
Bournemouth,Liverpool : L
Watford,Southampton : D
Everton,Liverpool : L
West Ham United,Watford : L
Brighton & Hove Albion,Aston Villa : D
Manchester City,Sheffield United : W
Manchester United,Aston Villa : D
Southampton,Brighton & Hove Albion : D
Watford,Crystal Palace : D
Southampton,Liverpool : L
Burnley,Norwich City : W
Aston Villa,West Ham United : D
```

```
Newcastle United,Sheffield United : na
West Ham United,Brighton & Hove Albion : D
Brighton & Hove Albion,Sheffield United : L
Everton,Watford : W
Bournemouth,Watford : L
Everton,Arsenal : D
Aston Villa,Chelsea : na
Liverpool,Chelsea : W
Arsenal,Sheffield United : D
Bournemouth,Leicester City : W
Everton,Burnley : W
Manchester United,Tottenham Hotspur : W
Tottenham Hotspur,Watford : D
Crystal Palace,Newcastle United : W
Wolverhampton Wanderers,Leicester City : D
Chelsea,Leicester City : D
Leicester City,Crystal Palace : L
Crystal Palace,Wolverhampton Wanderers : D
Leicester City,Bournemouth : W
Tottenham Hotspur,Everton : D
Liverpool,Southampton : W
Wolverhampton Wanderers,Crystal Palace : L
Aston Villa,Sheffield United : na
Watford,Manchester United : W
Newcastle United,Burnley : D
Wolverhampton Wanderers,Chelsea : L
Liverpool,West Ham United : W
Newcastle United,Arsenal : L
Bournemouth,Wolverhampton Wanderers : L
Burnley,West Ham United : W
Chelsea,Bournemouth : L
Liverpool,Wolverhampton Wanderers : W
Newcastle United,Leicester City : L
Burnley,Wolverhampton Wanderers : W
Watford,Aston Villa : W
Manchester United,Leicester City : W
Wolverhampton Wanderers,Manchester City : W
West Ham United,Manchester United : W
Brighton & Hove Albion,West Ham United : D
Crystal Palace,Tottenham Hotspur : L
Newcastle United,Everton : L
Everton,West Ham United : W
Manchester City,Everton : W
West Ham United,Aston Villa : na
West Ham United,Bournemouth : W
Burnley,Leicester City : W
Norwich City,Everton : na
Everton,Leicester City : L
```

List of teams with every other team and output the 38 team list

```
####################################
EPL Team Rankings are as below
####################################
Liverpool   = 6.491732145955493
Manchester City  = 6.111346543038106
Manchester United  = 6.050197146654108
Arsenal  = 5.920870352561915
Leicester City  = 5.859720956177918
Chelsea  = 5.825632257323818
Everton  = 5.825065479288827
Wolverhampton Wanderers  = 5.780116552390002
Brighton & Hove Albion  = 5.780116552390001
Tottenham Hotspur  = 5.780116552390001
Burnley  = 5.716863813320775
Bournemouth  = 5.643752916026365
Watford  = 5.63515606684763
Sheffield United  = 5.608424072843742
Crystal Palace  = 5.598298370571818
Newcastle United  = 5.547034960181148
West Ham United  = 5.444113098336445
Southampton  = 5.383530479987442
Aston Villa  = 5.336189087514663
Norwich City  = 5.324286748779795


####################################
Probability of given teams
####################################
Probability of : Liverpool 0.75
Probability of : Arsenal 0.25

Process finished with exit code 0
```

The final ranking along with probability of user selected teams

As the write up provided by professor states – Although Liverpool must end the season at the top of the table (it is mathematically impossible for any other team to pass them) Thus for summer we have Manchester City, Manchester United, Arsenal and Leicester City making the top five for this current season. And the teams being relegated are Southampton, Aston Villa and Norwich City this summer.

6. Conclusion

Thus for the season of 2018-2019 the rankings are concluded as below -

Rank	Team	Ratings
1	Liverpool	6.727272727272757
2	Manchester City	6.727272727272757
3	Arsenal	6.500000000000027
4	Chelsea	6.454545454545486
5	Manchester United	6.272727272727298
6	Tottenham Hotspur	6.272727272727298
7	Everton	6.181818
8	Wolverhampton Wanderers	6.181818
9	West Ham United	6.090909090909119
10	Bournemouth	6.045454545454578
11	Watford	5.954545454545484
12	Leicester City	5.954545454545484
13	Southampton	5.909090909090938
14	Newcastle United	5.863636363636381
15	Brighton & Hove Albion	5.863636363636381
16	Burnley	5.818181818181837
17	Fulham	5.772727272727298
18	Crystal Palace	5.772727272727298
19	Cardiff City	5.727272727272761
20	Huddersfield Town	5.4090909090909305

& for the current season the results are

Rank	Team	Ratings
1	Liverpool	6.491732145955493
2	Manchester City	6.111346543038106
3	Manchester United	6.050197146654108
4	Arsenal	5.920870352561915
5	Leicester City	5.859720956177918
6	Chelsea	5.825632257323818
7	Everton	5.825065479288827
8	Wolverhampton Wanderers	5.780116552390002
9	Brighton & Hove Albion	5.780116552390001
10	Tottenham Hotspur	5.780116552390001
11	Burnley	5.716863813320775
12	Bournemouth	5.643752916026365
13	Watford	5.63515606684763
14	Sheffield United	5.608424072843742
15	Crystal Palace	5.598298370571818
16	Newcastle United	5.547034960181148
17	West Ham United	5.444113098336445

18	Southampton	5.383530479987442
19	Aston Villa	5.336189087514663
20	Norwich City	5.324286748779795

6. References

1. https://towardsdatascience.com/generate-sports-rankings-with-data-science-4dd1979571da

2.
https://www.colleyrankings.com/matrhttps://www.colleyrankings.com/matrate.pdfate.pdf

3. Dataset for 2019-2020
https://en.wikipedia.org/wiki/2019%E2%80%9320_Premier_League

4. Dataset for 2018-2019
https://en.wikipedia.org/wiki/2019%E2%80%9320_Premier_League